



Microsoft Cognitive Services

Ms Rupish Arora¹, Jigyasa Sharma², Urja Mali³, Akanaksha Sharma⁴, Priya Raina⁵

Assistant Professor¹, Student^{2,3,4,5}

Department Of Computer Science Engineering

Mahant Bachittar Singh College of Engineering and Technology, J&K, India

Abstract:

Microsoft Cognitive Services are collection of machine learning algorithm which helps in solving various problems in the field of artificial intelligence, like language processing, machine learning search, computer vision etc. Basically Cognitive Services are collection of APIs, SDKs and services designed for developers. These services can make the applications more intelligent and more interactive. The aim of these services is to supply interesting and well-off computing experience. The available APIs of Microsoft Cognitive Services are Language API, Vision API, Speech API, Knowledge API, etc. Each API performs different functions such as language API identifies and discovers the requirements of the user, vision API examines the images and videos for useful information, speech API helps in identifying the speaker and knowledge API captures research from scientific account. By using these APIs developers can add the intelligent features like understanding face detection, speech detection, vision detection and recognition, emotion detection and video detection. The characteristics which distinguish Microsoft Cognitive Services from other services are multiple face tracking in less time, more accuracy in face recognition, presence of emotions with their types and percentages, better APIs. Microsoft Cognitive Service APIs are used in various fields like enhancing the security, expressing farcical moments, engaging customers via chat, etc.

Keywords: Knowledge API, Language API, Microsoft Cognitive Services, Speech API, Vision API.

I. INTRODUCTION

Cognitive Services are basically a set of machine learning algorithms designed by Microsoft in order to solve the problems in the field of Artificial Intelligence (AI) [1]. Cognitive Services provides the best suite for building an automated, interactive model of an application. It helps in constructing and adding powerful intelligence into the applications to allow natural and contextual interactions [2]. Microsoft Cognitive Services allow us to design applications with powerful algorithms using just a few lines of code. Microsoft's aim with these services is to provide more personal and rich computing experiences [2]. Cognitive Services come with a series of APIs and SDKs which are used to allow natural and contextual interaction within the applications that the developer is developing [1]. Microsoft Cognitive Services are of various types and they are grouped under five main categories: language, speech, vision, search and knowledge API. The goal of cognitive services is to balance AI by packaging it into distinct components those are easy to use in the developer's applications. Microsoft Cognitive Services are more related with expansion of Microsoft machine learning APIs, which can be used by developers for adding rich and intelligent features in their applications like video detection, face detection, speech, vision recognition, identifying and understanding emotions of the users and language understanding [3]. These services mainly target on duplicating a desired behaviour that a particular application demands without the restraints of cognitive plausibility [3].

II. COGNITIVE SERVICE APIS

Microsoft Cognitive Services offers various APIs and these can be used individually or can be combined. These APIs are:

1) Language API

Language API allows applications to process natural language, evaluate topics, emotions and beliefs and tells how to identify

and recognize the needs of the user [3]. There are various APIs that come under this. They are:

• Text Analytics API

It is a suite of text analytics web services constructed with Microsoft's machine learning. This API evaluates emotions beliefs and topics to understand the command of a user. The API is constructed to detect the key phrases, sentiments, topics, and language from our text. We don't require any training data to use this API, just bring your text data. This API uses advanced natural language processing techniques to give best in class predictions [4].

• Web Language Model API

This API automates a variety of standard, natural language processing tasks like state-of-the-art language modelling APIs. **This API supports following lookup operations:** Word breaking: Inserts spaces into string of words that do not have spaces, such as a hashtag or a part of a URL.

Predictive text: Gets a list of suggestions that are likely to follow the given words.

• Language Understanding Intelligent Service (LUIS)

This API helps an application to understand the commands given by the users. This API is designed to accommodate us with a way to create models easily, which in turn allows our applications to understand the commands of the user [5].

• Translator text API

This API helps to build new applications, tools, websites, or multi-language support requiring solutions. This API is an automatic transition serviced which supports multiple languages. It is a cloud based service which easily administers real-time text translation [4].

• Bing spell check API

This API basically detects and then corrects spelling mistakes in our applications. This API corrects spelling errors for users, identify the differences among the brand names, slang, names, etc and understand homonyms as the user types them. It provides proofreading capability as well [5].

2) Vision API

Vision API helps to build most personalized applications by returning smart insights about images, faces and emotions.

• Face API

The Microsoft Face API provides the most mature algorithm. This API performs functions such as detecting face attributes and recognizing the face. It basically helps in detecting, analyzing and organizing in given images. This API detects human face with high precision. It allows us to tag faces in any given photo. Alternatively, face detection allows extraction of face attributes like gender, age, facial hair, glasses, head pose etc. This API provides four face recognition functions: face verification, finding similar faces, face grouping, and identification. This API helps in adding super cool intelligence while building the applications [6].

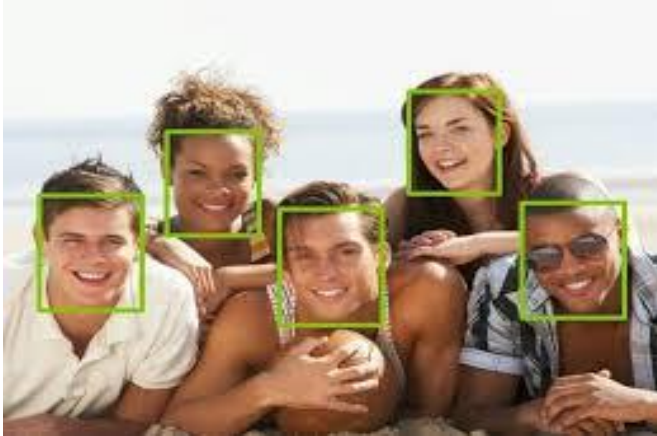


Figure.1. Face API

• Emotion API

Microsoft Emotion API helps in building applications which recognizes feelings, responds to various moods on the human face, and helps in getting closer to the user. This cloud based API helps in building personalized applications with edge-cloud based emotion recognition algorithm. This API, using the face based attributes, can detect happiness, sadness, anger, fear, and surprise. The algorithms used in this API detect the emotions according to universal facial expressions [7].



Figure.2. Emotion API

• Computer Vision API

Microsoft Computer Vision API takes selective information from the images to organize and process the data and also protects the user from undesired contents. This cloud based

API uses computer vision algorithms which helps in analyzing the visual contents in different ways:

- **Analysing:** Extracting information found in image
- **Generate thumbnails:** Generate high quality thumbnails as per the input image.
- **Read text in the image:** Allows taking photos of the text and helps in saving time and efforts [8].

3) Speech API

This API helps to improve speech recognition and identify the speaker. It basically is about processing the language spoken by the user in our applications. This includes various APIs categorized under this are:

• Bing speech API

Bing Speech API allows speech capabilities on internet connected devices. It provides speech-to-text, text-to-speech and language understanding capabilities dispatched via the cloud. Every standard platform, such as IOS, Windows, Android, and third party IoT devices are backed under this API. Microsoft uses this API for application like Cortana, Bing Torque [9]. It provides:

Text to speech conversion: This is used when the applications want to talk back with their users.

Speech Intent Recognition: It anatomizes the intent of speaker in order to generate actions with in the applications.

Speech Recognition: It converts the audio spoken by the user into text in real time.

• Speaker recognition API

This speaker recognition API makes use of speech to identify and authenticate each speaker. It helps our application to identify who is talking.

Speaker verification: This is a feature of voice authentication.

Speaker identification: This is a feature matches and compares the speech of known speakers from a group. It matches speech with its speaker [10].

• Translator speech API

Translator speech API is an automatic translation service. This API is cloud based and helps the developers to attach real-time, end-to-end, speech translations to our applications. It provides speech translation to our applications making it effective for real life conversation [11].

4) Knowledge API

The Knowledge API matches the complex information in order to resolve recommendation and semantic search used cases. This API allows us to link our grouped knowledge and helps in accessing that knowledge in quite effective ways.

• Recommendation API

Microsoft Recommendation API helps in providing personalized recommendation for the users. The API is made flexible enough to provide assistance for adding new features by third-party researchers or developers.

• Academic knowledge API

Microsoft academic knowledge API gathers user's queries for academic content in the Microsoft academic graph. This API examines relation among academic paper general and authors [12].

III. WHY COGNITIVE SERVICES ARE BETTER?

Companies like Microsoft Cognitive Services take big technologies and are good at building services for them. There are some incredible machine learning and artificial intelligence platforms, libraries and APIs that can be used for those technologies. Recently Microsoft has launched various services under the banner of cognitive services and they are quite impressive and are better than previous services like

Amazon, IBM, Google, Open CV, and Affectiva in the following ways:

Table.1. Comparison between cognitive services and other services

	Amazon	Google	Microsoft	IBM	Affectiva	Open CV
Face Detection	Yes	Yes	Yes	Yes	Yes	Yes
Face recognition (image)	Yes	No	Yes	No	No	No
Emotional depth (%)	No	No	Yes	No	Yes	No
Emotions present (Y/N)	Yes	Yes	Yes	No	Yes	No
Multiple-face tracking	Yes	Yes	Yes	Yes	Yes	No
API	Yes	Yes	Yes	Yes	Yes	No
Age and Gender	Yes	No	Yes	Yes	Yes	No

- Microsoft Cognitive Services provides much better and more accurate image face recognition as compare to Amazon whereas IBM, Affectiva, open CV, and Google do not provide this feature at all.
- Microsoft Cognitive Services provides more emotional depth percentage than affectiva while Amazon, Open CV, Google, IBM do not provide this function at all.
- Microsoft Cognitive Services allows multiple face tracking in very less time as compare to Amazon, Google, Open CV, and IBM.
- APIs given by Microsoft Cognitive Services are way better than the APIs provided by Google, Amazon, Affective and IBM while open CV does not use APIs.
- Microsoft Cognitive Services mentions the presence of emotions and its types along with its percentage whereas Open CV and IBM do not have this feature [5].

IV. APPLICATIONS

- **Enhancing the security**
The Microsoft face API allows verification of an image for acute authentication. This API helps in building the application with the ability to compare the images with varying attributes like pose, lightening, focus. This API allows correct face matching even in variety of conditions. Hence it allows face identification and verification at very fast speed without slowing the user down [13].
- **Expressing farcical moments immediately**
The Microsoft emotion API allows quick retrieval of emotions like happy, sad, and surprised out of millions by using multiple other APIs in combination. This API can search

through video frames to extract the right moment for the required content [5].



Figure .3. Expressing farcical moments

- **Enabling voice interaction with speech**
Microsoft’s speech API allows building of voice interactions between the application and user. Speech API enhances the speech recognition using complex technical word training and background noise reduction [14].
- **Engaging customers via chat**
Microsoft Cognitive Service API working along with some special frameworks help in engaging users on a whole new level. This allows the user to ask questions and escalating to the human operator if required [15].

V. CONCLUSION

The Cognitive Services are very up-and-coming, assuring and sum of these services are already being used in various application areas. The developers use these services to improve the applications by adding some new, rich and intelligent features. These services basically aim at providing more personal and rich computing experiences for users as well as developers. The range of application areas of these cognitive services is less but it can be expanded by using their services along with other modern services and techniques [16].

VI. REFERENCES

[1].L.A.B. Prieto and Z.K. Oplatkova. (2017). “Comparing the performance of Emotion-Recognition Implementations in OpenCV, Cognitive Services, and Google Vision APIs.” *WSEAS transactions on information science and applications*. E-ISSN: 2224-3402, Vol. 14.

[2].Eddie, C. L., Chan, G. B., and Mak, S. C. (2010). “Cognitive Location-Aware Information Retrieval by Agent-based Semantic Matching.” *International Journal of Cognitive Informatics & Natural Intelligence*.

[3].M. Coccoli, P. Maresca, and L. Stanganelli. (2016). “Cognitive computing in education.” *Journal of E-Learning and Knowledge Society*. ISSN: 1826-6223, e-ISSN: 1971-8829, Vol. 12.

[4].A. Moreno and T. Redondo. “Text Analytics: the convergence of Big Data and Artificial Intelligence.”

- [5]. Mohammed, Abdul Adeel, Rashid Minhas, QM Jonathan Wu, and Maher A. Sid-Ahmed. (2011). "Human face recognition based on Multidimensional PCA and extreme learning machine." *Pattern Recognition* 44, no. 10: 2588-2597.
- [6]. Vinay Aa, Vinay S Shekhara, Rituparna Jb, Tushar Aggrawalb, K N Balasubramanya Murthya, S Natarajanb. (2015). "Cloud Based Big Data Analytics Framework for Face Recognition in Social Networks using Machine Learning." *2nd International Symposium on Big Data and Cloud Computing (ISBCC'15)*.
- [7]. K.V. Bhuvneshwari, A. Abirami, and N. Sripriya. (2017, April 4). "Face Recognition using PCA." *International Journal of Engineering and Computer Science*. ISSN: 2319-242. Vol. 6, pp. 20820-20826.
- [8]. J. Weng and M. Luciw. (2010). "Online learning for attention, recognition, and tracking by a single developmental Framework." *Proceedings of the Conference on Computer Vision and Pattern Recognition*.
- [9]. S. Furui. (2005, November). "50 years of Progress in speech and Speaker Recognition Research." *ECTI Transactions on Computer and Information Technology*. Vol.1. No.2.
- [10]. H.Sakoe and S.Chiba. (1978, February). "Dynamic Programming Algorithm Optimization for Spoken Word Recognition." *IEEE Trans.Acoustics, Speech, Signal Proc.*
- [11]. M. A. Anusuya and S. K. Katti. (2009). "Speech Recognition by Machine." *International Journal of Computer Science and Information Security (IJCSIS)*, Vol. 6, No. 3.
- [12]. S. E. Hug, M. Ochsner, and M. P. Brandle. (2016). "Citation Analysis with Microsoft Academic." *Scientometrics*.
- [13]. Ramanathan, Narayanan, R. Chellappa, and A. K. Roy-Chowdhury. (2004). "Facial similarity across age, disguise, illumination and pose." In *Image Processing, 2004. ICIP'04. 2004 International Conference*. Vol. 3, pp. 1999-2002. IEEE.
- [14]. A. Berthold and A. Jameson. (1999). "Interpreting Symptoms of Cognitive Load in Speech Input." *User modeling: Proceedings of the seventh international conference*. pp. 235-244.
- [15]. S. L. Epstein, Rebecca -Passonneau, J. Gordon, and T. Ligorio. (2012). "The Role of Knowledge and Certainty in Understanding for Dialogue." *AAAI Fall Symposium: Advances in Cognitive Systems*.
- [16]. C. Eliasmith and C. Kolbeck. (2015). "Topics in Cognitive Science." *Marr's Attacks: On Reductionism and Vagueness*. Pages 1-13.